Asymptotic laws for regenerative compositions: gamma subordinators and the like *

Alexander Gnedin †, Jim Pitman ‡ and Marc Yor §

February 1, 2008

Abstract

For $\widetilde{\mathcal{R}} = 1 - \exp(-\mathcal{R})$ a random closed set obtained by exponential transformation of the closed range $\mathcal{R}$ of a subordinator, a regenerative composition of generic positive integer n is defined by recording the sizes of clusters of n uniform random points as they are separated by the points of $\widetilde{\mathcal{R}}$. We focus on the number of parts K_n of the composition when $\widetilde{\mathcal{R}}$ is derived from a gamma subordinator. We prove logarithmic asymptotics of the moments and central limit theorems for K_n and other functionals of the composition such as the number of singletons, doubletons, etc. This study complements our previous work on asymptotics of these functionals when the tail of the Lévy measure is regularly varying at $0+$.

AMS 2000 *subject classifications.* Primary 60G09, 60C05.
Keywords: composition, regenerative set, occupancy problem, logarithmic singularity

*Research supported in part by N.S.F. Grant DMS-0071448

†Utrecht University; e-mail gnedin@math.uu.nl
‡University of California, Berkeley; e-mail pitman@stat.Berkeley.EDU
§University of Paris VI

1 Introduction

For a drift-free subordinator $(S_t, t \geq 0)$ define $\widetilde{\mathcal{R}} \subset [0,1]$ to be the range of the *multiplicative subordinator* $\widetilde{S}_t = 1 - \exp(-S_t)\,, t \geq 0$, obtained by the exponential transform. The *gaps* in $\widetilde{\mathcal{R}}$ are the interval components of the open set $[0,1] \setminus \widetilde{\mathcal{R}}$. The following construction of random compositions of integers has been studied in [7, 8, 9, 10, 13]. For each n, let $u_1, \ldots, u_n$ be an independent sample from the uniform distribution on $[0,1]$, also independent of $\widetilde{\mathcal{R}}$. Define an ordered partition of the sample into nonempty blocks by assigning two sample points u_i and u_j to the same block if and only if the points fall in the same gap in $\widetilde{\mathcal{R}}$. A composition $\mathcal{C}_n$ of integer n is defined by the sequence of sizes of blocks, ordered from left to right. The sequence of compositions $(\mathcal{C}_n)$ is a regenerative composition structure in the sense of [9]. The regeneration property means that, conditionally given the first part of $\mathcal{C}_n$ is r, the remaining composition of $n-r$ has the same distribution as $\mathcal{C}_{n-r}$.

The purpose of this paper is to prove central limit theorems and determine the asymptotic behaviour of moments for the number of parts K_n and some other functionals of the composition $\mathcal{C}_n$. We focus on the gamma subordinator with parameter $\theta > 0$, that is with Lévy measure $\nu(\mathrm{d}y) = (e^{-\theta y}/y)\,\mathrm{d}y\,$,. More generally, we obtain similar results for Lévy measures which are sufficiently like the gamma Lévy measure in their asymptotic behaviour at both $0+$ and $\infty-$. In the case of gamma subordinators our principal result (Theorem 11) specialises as follows:

Theorem 1 *For the gamma subordinator with parameter $\theta > 0$ the number of parts K_n is asymptotically normal, with $\mathbb{E}\,K_n \sim (\theta/2)\log^2 n$ and $\operatorname{var} K_n \sim (\theta/3)\log^3 n$.*

Other variables under consideration are $K_n(t)$, the number of parts of a partial composition produced by the subordinator restricted to $[0,t]$, and the counts $K_{n,r}$ and $K_{n,r}(t)$ defined as the multiplicity of part r in $\mathcal{C}_n$ and the multiplicity of part r in the partial composition, respectively.

In our previous paper [10] we studied K_n for a subordinator whose Lévy measure is regularly varying at 0 with index $\alpha \in]0,1]$. In that case we obtained limit theorems for K_n with normalisation by $n^\alpha\,\ell(n)$ where ℓ is a function of slow variation at ∞. According to these results, $K_n/(n^\alpha\,\ell(n))$ converges strongly and with all moments to a nondegenerate limit, which is not gaussian. Moreover, $K_n(t)$ is of the same order of magnitude as K_n, for each $t > 0$. For the case of a compound Poisson process, when the Lévy measure is finite, Gnedin [7] had previously shown that K_n typically exhibits a normal limit with both expectation and variance of the order $\log n$, though $K_n(t)$ remains bounded as n grows, for each $t > 0$. Thus the case of gamma-type subordinators may be seen as intermediate between the above two: as in the regular variation case, $K_n(t)$ is unbounded, but, as in the case of a finite Lévy measure, its contribution to K_n is asymptotically negligible, for each $t > 0$.

Results akin to Theorem 1 are very different from the limit theorems in [7, 10] and require other techniques. Our approach here relies on a recent version of the contraction method due to Neininger and Rüschendorf [12]. Application of this method requires an appropriate decomposition of K_n, control of the principal asymptotics of two moments, and an estimate of the remainder term in the asymptotic expansion of the variance. We provide this background by poissonising $\mathcal{C}_n$ and applying the Mellin transform technique to analyse integral recursions for the moments.

We also establish joint convergence to a gaussian limit for the sequence of small-part counts $(K_{n,r}, r \geq 1)$, as $n \to \infty$. This kind of convergence holds neither in the regular variation case (when the scaled $K_{n,r}$'s converge to multiples of the same variable) nor in the compound Poisson case (when $(K_{n,r}, n \geq 1)$ is bounded uniformly in n, for each fixed r). It resembles, however, Karlin's [11] central limit theorem for nonrandom frequencies in the case of regular variation with a positive index.

2 Setup and notation

Each *part* of $\mathcal{C}_n$ corresponds to a nonempty cluster of the uniform sample points within some gap. So K_n is the number of gaps occupied by at least one of n sample points, and $K_{n,r}$ is the number of gaps that contain exactly r sample points. Keep in mind that $K_n = \Sigma_r\,K_{n,r}$, and $n = \Sigma_r\,rK_{n,r}$. Sometimes K_n is called the *length*, and n the *weight* of the composition. Similarly, $K_n(t)$ and $K_{n,r}(t)$ are the counts of

sizes of clusters within the gaps of a smaller set $\widetilde{\mathcal{R}} \cap [0, \widetilde{S}_t]$, corresponding to the subordinator restricted to $[0,t]$.

For ν the Lévy measure of (S_t) let $\widetilde{\nu}$ be the measure on $[0,1]$ obtained from ν by the exponential transform $y \mapsto 1-e^{-y}$. The Laplace exponent of the subordinator is given by the formulas

$$\Phi(s) = \int_0^\infty (1-e^{-sy})\nu(\mathrm{d}y) = \int_0^1 (1-(1-x)^s)\widetilde{\nu}(\mathrm{d}x)\,, \qquad \Re\, s \geq 0\,. \tag{1}$$

Considering also the binomial moments

$$\Phi(n:m) := \int_0^1 \binom{n}{m} x^m (1-x)^{n-m} \widetilde{\nu}(\mathrm{d}x)\,, \qquad 1 \leq m \leq n\,, \tag{2}$$

the ratio $\Phi(n:m)/\Phi(n)\,, m = 1,\dots,n$, gives the distribution of the first part of $\mathcal{C}_n$, and the probability of a particular value of $\mathcal{C}_n$ is a product of factors of this type [9]. We also record the relation between power and logarithmic moments of measures ν and $\widetilde{\nu}$:

$$\mathtt{m}_j := \int_0^\infty y^j\,\nu(\mathrm{d}y) = \int_0^1 |\log(1-x)|^j\,\widetilde{\nu}(\mathrm{d}x)\,,\ j = 1,2,\dots\,. \tag{3}$$

As in [10], we shall also consider the *poissonised* composition $\widehat{\mathcal{C}}_\rho$ derived from $\widetilde{\mathcal{R}}$ by the same construction as $\mathcal{C}_n$, but with the set of atoms of a homogeneous Poisson point process on $[0,1]$ with rate ρ, instead of the uniform sample of fixed size. We denote $\widehat{K}_\rho$, $\widehat{K}_{\rho,r}$, $\widehat{K}_\rho(t)$, $\widehat{K}_{\rho,r}(t)$ the obvious counterparts of the fixed-n variables, and introduce the Poisson transform of the Laplace exponent

$$\widehat{\Phi}(s) := \int_0^1 (1-e^{-sx})\widetilde{\nu}(\mathrm{d}x) \;=\; e^{-s}\sum_{n=0}^\infty \frac{s^n}{n!}\,\Phi(n). \tag{4}$$

We shall be using throughout the abbreviations

$$\mathrm{L} = \log\rho, \qquad \log^a x = (\log x)^a\,.$$

Our primary concern are the compositions induced by a *gamma subordinator*, whose Lévy measure, its exponential transform and the Laplace exponent are given in terms of a parameter $\theta > 0$ by

$$\nu(\mathrm{d}y) = \frac{e^{-\theta y}}{y}\,\mathrm{d}y\,,\ y \in]0,\infty[\,, \quad \widetilde{\nu}(\mathrm{d}x) = \frac{(1-x)^{\theta-1}}{|\log(1-x)|}\,\mathrm{d}x\,,\ x \in]0,1]\,. \quad \Phi(s) = \log\left(1+\frac{s}{\theta}\right)\,, \tag{5}$$

The density of $\widetilde{\nu}$ has a pole and the tail $\widetilde{\nu}[x,1]$ has a logarithmic singularity at 0. We could consider a larger family of measures which differ from (5) by a positive factor, but this would not increase generality, since multiplying the Lévy measure by such a factor does not affect the laws of $\mathcal{R}$, $\mathcal{C}_n$, K_n or $\widehat{K}_\rho$ (although it requires a linear time-change for functionals of a partial composition like $K_n(t)$ or $\widehat{K}_\rho(t)$).

More generally, we consider a drift-free subordinator with Lévy measure ν, which has a continuous density on $]0,\infty[$ and satisfies the following conditions (L) and (R) which we record in terms of ν and $\widetilde{\nu}$:

(L) either of the following four equivalent conditions holds

$$\Phi(\rho) = \mathrm{L} + c + O(\rho^{-\epsilon})\,,\ \text{as } \rho\to\infty\,, \qquad \widehat{\Phi}(\rho) = \mathrm{L} + c + O(\rho^{-\epsilon})\,,\ \text{as } \rho\to\infty$$

$$\nu[y,\infty] = -\log y + c - \gamma + O(y^\epsilon)\,,\ \text{as } y\downarrow 0\,, \qquad \widetilde{\nu}[x,1] = -\log x + c - \gamma + O(x^\epsilon)\,,\ \text{as } x\downarrow 0$$

where c and $\epsilon > 0$ are some constants and $\gamma = -\Gamma'(1)$ is the Euler constant,

(R) either of the following two equivalent conditions holds

$$\nu[y,\infty] = O(e^{-\epsilon y})\,,\ y\uparrow\infty\,, \qquad \widetilde{\nu}[x,1] = O((1-x)^\epsilon)\,,\ x\uparrow 1$$

where $\epsilon > 0$.

See Appendix for the equivalence of conditions in (L) (L stands for *left* and *logarithmic* at 0). Condition (R) (R for *right* or *regular at* 1) implies that all moments $\mathtt{m}_j$ are finite, and that Φ is analytical for $\Re s > -\epsilon$.

Throughout in this paper, ϵ or δ are some sufficiently small positive constants whose values are context-related and may change from line to line. We denote c the constant in (L). For a generic positive constant we write d, and use c_j, d_j for further real constants which are not important, but may appear with explicit evaluation in asymptotic expansions or other formulas. For shorthand, the right tail of $\widetilde{\nu}$ is denoted by

$$\vec{\nu}(x) = \widetilde{\nu}[x, 1]\,, \qquad x \in]0, 1]$$

and a homogeneous Poisson point process with intensity ρ on $[0, 1]$ is denoted PPP(ρ).

3 Recursions

We shall make use of two types of recursions. These apply to regenerative compositions generated by a drift-free subordinator, without any special assumptions on the Lévy measure. (They can also be readily generalised to the case with drift). The first type of recursion, suited to the Poisson framework, is based on a decomposition of $\widetilde{\mathcal{R}}$ analogous to the first-jump decomposition of a renewal process. The second type is based on splitting the range of a subordinator by its value at a stopping time.

Integral recursions Define a *pattern* E to be a nonempty set of positive integers. We say that a finite configuration of points within a given subinterval $]a, b[\, \in\,]0, 1[$ fits in E if the cardinality of the configuration is in E. For E fixed, let

$$\pi(\rho) = \sum_{r \in E} e^{-\rho} \rho^r / r!$$

be the probability that the configuration of atoms of PPP(ρ) fits in the pattern E. The probability that the PPP(ρ) configuration on a subinterval $]a, b[$ fits in E is equal to $\pi(\rho(b - a))$.

Consider the poissonised composition induced by an arbitrary drift-free subordinator with $\widetilde{\nu}\{1\} = 0$ (no killing). For E fixed, let N_ρ be the number of gaps of $\widetilde{\mathcal{R}}$ such that the PPP(ρ) configuration within the gaps fits in E. The count N_ρ is a functional of the poissonised composition $\widehat{\mathcal{C}}_\rho$; in particular, defining E to be $\{1, 2, \ldots\}$ (the configuration is nonempty) or $E = \{r\}$ (the configuration consists of r points) we obtain $N_\rho = \widehat{K}_\rho$, respectively $N_\rho = \widehat{K}_{\rho,r}$. Other possibilities may be considered, for example taking $E = \{1, 3, 5, \ldots\}$ the count N_ρ becomes the number of odd parts of the composition.

Let $p_j(\rho) = \mathbb{P}(N_\rho = j)$ be the distribution of N_ρ for some fixed pattern. Each p_j may be extended to an entire function of a complex variable, with the initial value $p_j(0) = 1(j = 0)$. Introduce the factorial moments

$$f^{(m)}(\rho) = \mathbb{E} N_\rho(N_\rho - 1) \cdots (N_\rho - m + 1)\,, \;\; m = 0, 1, \ldots$$

with the convention $f^{(0)}(\rho) = 1$. The following lemma is a minor variation of [10, Lemma 6.1].

Lemma 2 *Let E be a pattern with the probability of occurrence $\pi(\rho)$. The distribution of N_ρ satisfies the integral recursion*

$$\int_0^1 (p_j(\rho) - (1 - \pi(\rho x))\, p_j(\rho(1 - x))\, \widetilde{\nu}(\mathrm{d}x) \; = \int_0^1 \pi(\rho x) p_{j-1}(\rho(1 - x))\, \widetilde{\nu}(\mathrm{d}x) \tag{6}$$

for $j = 1, 2, \ldots$ and $\rho \geq 0$. The same equation is also valid for $j = 0$, with 0 on the right-hand side. Furthermore, the factorial moments of N_ρ satisfy the recursion

$$\int_0^1 (f^{(m)}(\rho) - f^{(m)}(\rho(1 - x))\, \widetilde{\nu}(\mathrm{d}x) \; = m \int_0^1 \pi(\rho x) f^{(m-1)}(\rho(1 - x))\, \widetilde{\nu}(\mathrm{d}x) \tag{7}$$

which taken together with $f^{(0)}(\rho) = 1$ and $f^{(m)}(0) = 1(m = 0)$ uniquely determines them.

Proof. The derivation of recursions follows as in [10, Lemma 6.1]. The uniqueness claim is a consequence of analyticity. $\square$

Integral recursions for the distribution and the factorial moments of $\widehat{K}_\rho$ follow by taking $\pi(\rho) = 1-e^{-\rho}$ which is the probability that the PPP(ρ) configuration is nonempty; we have then $p_0(\rho) = e^{-\rho}$. To obtain recursions for $\widehat{K}_{r,\rho}$ we should take $\pi(\rho) = e^{-\rho}\rho^r/r!$ (in this case no simple formula for $p_0(\rho)$ is known).

Splitting at an independent exponential time Further recursions follow by splitting the range of a subordinator by its value at a stopping time. Though the fixed-n version is needed to apply [12], we focus on the poissonised model, which simplifies moment computations. Transfer of results to the fixed n model then follows by elementary depoissonisation.

For each $t \geq 0$, consider the sigma-algebra generated by both $(\widetilde{S}_u, u \leq t)$ and the restriction of PPP(ρ) to $[0, \widetilde{S}_t]$. As t varies, this defines a filtration, and we can consider stopping times with respect to it. Let τ be such a stopping time with range $[0, \infty]$ and let $b_\rho = \widehat{K}_\rho(\tau)$ be the number of parts of $\mathcal{C}_\rho$ produced by the subordinator up to time τ. The strong Markov property of $\widetilde{\mathcal{R}}$ along with the independence property of PPP imply that the number of parts satisfies the distributional equation

$$\widehat{K}_\rho \stackrel{d}{=} b_\rho + \widehat{K}'_{I_\rho} \tag{8}$$

where $b_\rho = \widehat{K}_\rho(\tau)\,, I_\rho = \rho(1-\widetilde{S}_\tau)$ and $(\widehat{K}'_\rho)$ is a distributional copy of $(\widehat{K}_\rho)$, independent of (b_ρ, I_ρ). For example, letting τ be the first time that a jump of the subordinator covers at least one Poisson point, the equation holds with $b_\rho = 1$. Identities analogous to (8) can be written also for $\widehat{K}_{\rho,r}$ and more general pattern counts.

We shall consider in some detail the most important case when τ is an exponential time with rate λ, independent of the subordinator and the PPP (thus τ is a *randomised* stopping time with respect to the above filtration). The stopped process has an obvious interpretation as a *killed* multiplicative subordinator, which jumps at time τ to the terminal value 1 (thus producing the final *meander* gap in the range). In accord with our previous notation, let $\widehat{K}_\rho(\tau)$ be the number of parts produced by the subordinator *before* τ (thus a possible part induced by the meander gap is not taken into account). Let $(p_j(\rho), j \geq 0)$ and $(f^{(m)}, m \geq 0)$ be the distribution and factorial moments of $\widehat{K}_\rho(\tau)$.

Lemma 3 *For τ an exponential time with rate λ, independent of the process $(\widehat{K}_\rho(t), t \geq 0)$, the factorial moments of $\widehat{K}_\rho(\tau)$ satisfy the recursion*

$$\lambda\, f^{(m)}(\rho) + \int_0^1 (f^{(m)}(\rho) - f^{(m)}(\rho(1-x)))\, \widetilde{\nu}(\mathrm{d}x) \;=m \int_0^1 \pi(\rho x) f^{(m-1)}(\rho(1-x))\, \widetilde{\nu}(\mathrm{d}x) \tag{9}$$

which taken together with $f^{(0)}(\rho) = 1$ and $f^{(m)}(\rho) = 1(m=0)$ uniquely determines them.

Proof. In the case of a finite Lévy measure, the first-jump decomposition of the range yields

$$p_j(\rho) = \int_0^1 \Big((1-\pi(\rho x)) p_j(\rho(1-x)) + \pi(\rho x)\, p_{j-1}(\rho(1-x)) \Big) \frac{\widetilde{\nu}(\mathrm{d}x)}{\vec{\nu}(0) + \lambda}$$

$$p_0(\rho) = \frac{\lambda}{\vec{\nu}(0)+\lambda} + \int_0^1 (1-\pi(\rho x)) p_0(\rho(1-x)) \frac{\widetilde{\nu}(\mathrm{d}x)}{\vec{\nu}(0)+\lambda}\,.$$

To see the extension for an arbitrary Lévy measure, substitute

$$p_j(\rho) = p_j(\rho) \frac{\lambda}{\vec{\nu}(0)+\lambda} + \int_0^1 p_j(\rho) \frac{\widetilde{\nu}(\mathrm{d}x)}{\vec{\nu}(0)+\lambda}\,,$$

then rearrange terms and argue as in [10, Lemma 6.1]. $\square$

If the meander gap is counted (this corresponds to $\widehat{K}_\rho$ for killed subordinator) we should modify the recursion for moments by adding $\lambda\pi(\rho)1(m=1)$ to the right-hand side.

Remark. The distribution and moments of $\widehat{K}_\rho(t)$ may be obtained as the inverse Laplace transform in λ from the analogous quantities for $\widehat{K}_\rho(\tau)$.

4 Moments of $\widehat{K}_\rho$

Mellin transform resolution The recursion (7) is intrinsically related to a convolution-type integral equation

$$\int_0^1 (f(\rho) - f(\rho(1-x)))\widetilde{\nu}(\mathrm{d}x) = g(\rho) \tag{10}$$

with the function g given and the function f unknown. For g analytic, there is a unique analytic solution f with given initial value $f(0)$. Observe that constants d constitute the null space of the integral operator, thus for each solution f the function $f + d$ is another solution, with the initial value $f(0) + d$.

It is easy to write out a power series solution to (10), but such a representation itself does not help describe the large-ρ behaviour. We turn therefore to asymptotic methods based on the Mellin transform. Recall that for a locally integrable real-valued function ϕ on $[0, \infty]$ the Mellin transform is defined by the integral

$$M\phi(s) = \int_0^\infty \rho^{s-1}\phi(\rho)\,\mathrm{d}\rho$$

which is assumed to converge absolutely for s in some open interval on the real axis, hence also for s in the open vertical strip based on the interval. The left convergence abscissa of $M\phi$ is determined by the behaviour of ϕ near 0, while the right convergence abscissa is determined by the behaviour of ϕ at ∞.

The analysis to follow is based on the formula

$$Mf(s) = \frac{Mg(s)}{\Phi(-s)} \tag{11}$$

which is valid in the common domain of definition of all ingredients. The formula follows by Fubini and a change of variable from

$$\int_0^\infty \rho^{s-1}\mathrm{d}\rho \int_0^1 (f(\rho) - f(\rho(1-x))\widetilde{\nu}(\mathrm{d}x) = Mf(s)\int_0^1 (1-(1-x)^{-s})\widetilde{\nu}(\mathrm{d}x) = Mf(s)\Phi(-s)$$

according to formula (1). Another important tool is the following correspondence between asymptotic expansion of a function and singularities of its Mellin transform. See [5] for a fuller exposition of this technique.

Lemma 4 [5, Section 2] *Suppose the Mellin transform Mf of a function f is analytic in a strip $a < \Re s < b$. If Mf can be extended meromorphically through the right convergence abscissa in a larger trip $a < \Re s < b + \epsilon$ and has finitely many poles there, then each pole z and each term $d\,(s-z)^{-k-1}$ in the Laurent expansion of Mf at z contributes the term*

$$\frac{d\,(-1)^{k+1}}{k!}\,\rho^{-z}\,\log^k \rho$$

to the asymptotic expansion of f at ∞. The remainder term of the expansion of f is then $O(\rho^{-b-\epsilon})$. Conversely, the asymptotic expansion of f at ∞ with such a remainder implies the termwise singular expansion of Mf in the strip $a < \Re s < b + \epsilon$ provided that for some $\delta > 0$

$$|Mf(s)| = O(|s|^{-1-\delta})$$

as $|s| \to \infty$ in a strip $b' < \Re\, s < b + \epsilon$ for some $b' < b$.

For the expectation $f^{(1)}(\rho) := \mathbb{E}\,\widehat{K}_\rho$ the formulas (7) and (11) specialise as

$$Mf^{(1)}(s) = \frac{M\widehat{\Phi}(s)}{\Phi(-s)} \tag{12}$$

where $\widehat{\Phi}(s)$ is defined by (4). Using Fubini and integration by parts we transform the numerator as

$$M\widehat{\Phi}(s) = \int_0^\infty \rho^{s-1}\mathrm{d}\rho \int_0^1 \rho e^{-\rho x}\vec{\nu}(x)\,\mathrm{d}x = \int_0^1 \vec{\nu}(x)\,\mathrm{d}x \int_0^\infty \rho^s e^{-\rho x}\mathrm{d}\,\rho =$$
$$\Gamma(s+1)\int_0^1 x^{-s-1}\vec{\nu}(x)\mathrm{d}x = -\Gamma(s)\int_0^1 x^{-s}\,\widetilde{\nu}(\mathrm{d}x).$$

We can now re-write formula (12) as

$$Mf^{(1)}(s) = -\Gamma(s)\frac{\Phi(-s:-s)}{\Phi(-s)} = -\Gamma(s)\frac{\int_0^1 x^{-s-1}\,\vec{\nu}(x)\,\mathrm{d}x}{\int_0^1 (1-x)^{-s-1}\,\vec{\nu}(x)\,\mathrm{d}x}\,, \tag{13}$$

where we used (1) and (2) in the form

$$\Phi(-s:-s) = -s\int_0^1 x^{-s-1}\,\vec{\nu}(x)\,\mathrm{d}x\,, \qquad \Phi(-s) = -s\int_0^1 (1-x)^{-s-1}\,\vec{\nu}(x)\,\mathrm{d}x.$$

Expectation Assuming (L) we conclude from (13) that $Mf^{(1)}$ is defined in the strip $-1 < \Re s < 0$. We focus therefore on the meromorphic continuation of $Mf^{(1)}$ through the right convergence abscissa $\Re s = 0$. The same formula says that the poles of $Mf^{(1)}$ might be caused either by poles of Γ or by poles of $\Phi(-s:-s)/s$, or by zeros of $\Phi(-s)/s$. We shall consider the three ingredients separately.

For $\Re s > -1$ the gamma function has a unique pole at 0, with Laurent expansion

$$\Gamma(s) = s^{-1} - \gamma + d_1 s + O(s^2)$$

where $d_1 = \gamma^2/2 + \pi^2/12$. With reference to Stirling's approximation, when u is a bounded real number and v is a large real number

$$|\Gamma(u+\mathtt{i}v)| \sim (2\pi)^{1/2}|v|^{u-1/2}e^{-\pi|v|/2} \tag{14}$$

uniformly in u.

By assumption (R), the function

$$-\Phi(-s)/s = \int_0^1 (1-x)^{-s-1}\,\vec{\nu}(x)\,\mathrm{d}x$$

is analytic for $\Re s < \epsilon$, and its behaviour at complex infinity is regulated by the next lemma.

Lemma 5 *If ν has a continuous density on $]0,\infty[$ and satisfies* (L) *and* (R) *then*

$$\Phi(s) \sim \gamma\,\log|s|\,, \quad \text{as } |s| \to \infty$$

uniformly in any strip $-\epsilon < \Re s < d$, for ϵ sufficiently small.

Proof. The measure $y\,\nu(\mathrm{d}y)$ is finite, with exponential decay at infinity, thus for $s = u + \mathtt{i}v$ the partial derivative

$$\frac{\mathrm{d}}{\mathrm{d}u}\Phi(u+\mathtt{i}v) = \int_0^\infty e^{-uy}e^{-\mathtt{i}vy}y\,\nu(\mathrm{d}y)$$

is bounded for $u > -\epsilon$, uniformly in v. Hence $|\Phi(u+\mathtt{i}v) - \Phi(\mathtt{i}v)|$ is uniformly bounded both in v and in $u \in [-\epsilon, d]$. Thus it is sufficient to show the asymptotics for $u = 0$. But for the Fourier integral

$$\Phi(\mathrm{i}v) = \mathrm{i}v\int_0^\infty e^{-\mathrm{i}vy}\nu[y,\infty]\,\mathrm{d}y$$

the desired asymptotics follows from the expansion (L) and smoothness of $\nu[\,\cdot\,,\infty]$ by application of [4, Section 3.1, Theorem 1.11] (with a further reference to [1]). $\square$

By one further elementary lemma (see Appendix), condition (L) implies that $\Phi(-s)/s$ has no zeroes for $\Re\, s \le 0$, hence by Lemma 5 there are no zeroes in the strip $-1 < \Re\, s \le \epsilon$. The Taylor expansion at 0 involves the logarithmic moments (3)

$$\frac{-\Phi(-s)}{s} = \mathtt{m}_1 + \mathtt{m}_2\frac{s}{2!} + \mathtt{m}_3\frac{s^2}{3!} + O(s^3). \tag{15}$$

The integral

$$-\Phi(-s:-s)/s = \int_0^1 x^{-s-1}\vec{\nu}(x)\,\mathrm{d}x$$

converges for $\Re\, s < 0$ and by assumption (L) this function is meromorphic for $\Re\, s < \epsilon$ and in this half-plane has the unique pole at $s = 0$, where the Laurent series starts with

$$\frac{-\Phi(-s:-s)}{s} = s^{-2} - (c-\gamma)s^{-1} + d_2 + o(1)$$

where d_2 is some constant. The meromorphic extension is obtained from this formula, provided we substitute a suitable analytic function for the $o(1)$ term. This follows by computing

$$\int_0^1 (|\log x| + c - \gamma + h(x))x^{-s-1}\mathrm{d}x$$

with $h(x) = O(x^\epsilon)$ and $\Re\, s < 0$, and noting that the integral of h is analytic for $\Re\, s < \epsilon$ and it is bounded in this domain due to a maximum ridge on the real axis.

It follows that $Mf^{(1)}$ is meromorphic in the strip $-1 < \Re s < \epsilon$ with a unique singularity at 0, where it has a *triple* pole. Putting the ingredients together and with a minor assistance of `Mathematica` we obtain the Laurent expansion

$$Mf^{(1)}(s) = -\frac{1}{\mathtt{m}_1}s^{-3} + \left(\frac{c}{\mathtt{m}_1} + \frac{\mathtt{m}_2}{2\mathtt{m}_1^2}\right)s^{-2} + d_3 s^{-1} + O(1)$$

where

$$d_3 = \left(\frac{-d_1 - d_2 - c\gamma + \gamma^2}{\mathtt{m}_1} - \frac{c\mathtt{m}_2}{2\mathtt{m}_1^2} - \frac{\mathtt{m}_2^2}{4\mathtt{m}_1^3} + \frac{\mathtt{m}_3}{6\mathtt{m}_1{}^2}\right)$$

is a constant whose explicit value will not be used below. By Lemma 4, as $\rho \to \infty$, the asymptotic expansion of the expectation is

$$f^{(1)}(\rho) = \frac{1}{2\mathtt{m}_1}\mathrm{L}^2 + \left(\frac{\mathtt{m}_2}{2\mathtt{m}_1^2} + \frac{c}{\mathtt{m}_1}\right)\mathrm{L} - d_3 + O(\rho^{-\epsilon}). \tag{16}$$

The decay condition on $Mf^{(1)}$ required in Lemma 4 is satisfied, because this $Mf^{(1)}$ has exponential decay as $|s| \to \infty$ in a strip about the imaginary axis due to (14), Lemma 5 and because $|\Phi(s:s)| = O(|s|^{-1})$.

Evaluating the right-hand side of (7). We will use (11) once again, to compute $f^{(2)}$. The right-hand side of (7) is evaluated with the help of the next lemma.

Lemma 6 *Assume* (L) *and* (R) *and suppose f is an increasing positive function of $\rho \in [0,\infty]$, which admits an asymptotic expansion at ∞ as a polynomial in* L*, with a remainder $O(\rho^{-\epsilon})$. Then a similar expansion holds also for the function*

$$h(\rho) = \int_0^1 (1 - e^{-\rho x})f(\rho(1-x))\widetilde{\nu}(\mathrm{d}x)$$

and this expansion is obtained from that of f by replacing each L^k *with*

$$\mathrm{L}^{k+1} + c\mathrm{L}^k + \sum_{j=1}^{k} (-1)^j \binom{k}{j} \mathtt{m}_j \, \mathrm{L}^{k-j}$$

where c is as in (L).

Proof. By the assumption

$$f(\rho) = \sum_{j=0}^{k} d_j \, \mathrm{L}^j + O(\rho^{-\epsilon}). \tag{17}$$

Start with the case $f = \mathrm{L}^k$. The integral is computed by the binomial expansion of $(\mathrm{L} + \log(1-x))^k$ and termwise integration. The leading term does not depend on x and, in view of (L), it contributes $\mathrm{L}^k\,(\mathrm{L}+c) + O(\rho^{-\epsilon})$. The lower order terms are integrated using (3), with a remainder estimated by a constant multiple of

$$\mathrm{L}^k \int_0^1 e^{-\rho x} |\log(1-x)|^k \, \widetilde{\nu}(\mathrm{d}x)\,, \quad k \geq 1.$$

The last integral is evaluated as $O(\rho^{1-\epsilon})$ using integration by parts, using (L) and applying a standard Tauberian theorem on Laplace transforms.

By linearity we obtain asymptotic expansion for the *model integral* $I(\rho)$ which corresponds to the polynomial part of f, as in (17) but with zero remainder term. One checks easily using (R) that for $\delta := \rho^{-1/2}$ the contribution to $I(\rho)$ of the integral over $[1-\delta,\, 1]$ is estimated as $O(\rho^{-\epsilon})$.

For a general f as in (17) we again split the integral at $1-\delta$. For $x \in [0, 1-\delta]$ we can apply expansion (17) to $f(\rho(1-x))$ with a remainder $O(\rho^{-\epsilon})$, thus by (L) we have

$$\int_0^{1-\delta} (1 - e^{-\rho x}) f(\rho(1-x)) \widetilde{\nu}(\mathrm{d}x) = I(\rho) + O(\rho^{-\epsilon}).$$

It remains to show that the integral over $[1-\delta,\, 1]$ is $O(\rho^{-\epsilon})$ but this is easy because the assumed monotonicity yields the bound

$$\int_{1-\delta}^{1} (1 - e^{-\rho x}) f(\rho(1-x)) \widetilde{\nu}(\mathrm{d}x) < f(\rho\,\delta)\, \vec{\nu}(1-\delta)$$

and this has the desired order by (17), assumption (R) and our choice of δ. □

Taking $f^{(1)}$ for f, the lemma translates the expansion (16) into the $\rho \to \infty$ formula for the right-hand side of (7) with $m = 2$

$$g_1(\rho) = \frac{1}{\mathtt{m}_1} \mathrm{L}^3 + \left(\frac{3c}{\mathtt{m}_1} + \frac{\mathtt{m}_2}{\mathtt{m}_1^2} \right) \mathrm{L}^2 + d_4 \mathrm{L} + d_5 + O(\rho^{-\epsilon}) \tag{18}$$

where d_4, d_5 are some constants, and the factor 2 in (7) is included in the definition of g_1.

Moments of the second order Using the direct correspondence in Lemma 4, we see that 0 is the sole singularity of Mg_1, and derive the expansion at $s = 0$

$$Mg_1(s) = \frac{6}{\mathtt{m}_1} s^{-4} - \left(\frac{6c}{\mathtt{m}_1} + \frac{2\mathtt{m}_2}{\mathtt{m}_1^2} \right) s^{-3} + O(s^{-2})$$

which together with (11) and (15) implies

$$Mf^{(2)}(s) = \frac{Mg_1(s)}{\Phi(-s)} = -\frac{6}{\mathtt{m}_1^2}\, s^{-5} + \frac{6c\mathtt{m}_1 + 5\mathtt{m}_2}{\mathtt{m}_1^3}\, s^{-4} + O(s^{-3})$$

Finally, the inverse correspondence applied to $f^{(2)}$ yields

$$f^{(2)}(\rho) = \frac{1}{4\mathtt{m}_1^2}\mathrm{L}^4 + \left(\frac{c}{\mathtt{m}_1^2} + \frac{5\mathtt{m}_2}{6\mathtt{m}_1^3}\right)\mathrm{L}^3 + d\,\mathrm{L}^2 + O(\mathrm{L}) \tag{19}$$

provided the bound $|Mf^{(2)}(s)| = O(|s|^{-1-\epsilon})$ can be established. This requires some effort because for Mg_1 no explicit formula is available. Postponing justification of the decay condition, and recalling

$$\mathbb{E}\,\widehat{K}_\rho = f^{(1)}(\rho)\,, \quad \mathrm{var}\,\widehat{K}_\rho = f^{(2)}(\rho) + f^{(1)}(\rho) - (f^{(1)}(\rho))^2$$

we have from (16) and (19)

Theorem 7 *If the Lévy measure of subordinator has a continuous density on* $]0,\infty[$ *and satisfies the assumptions* (L) *and* (R)*, then as* $\rho \to \infty$ *with* $\mathrm{L} = \log\rho$*, the asymptotic expansions of the first two central moments of* $\widehat{K}_\rho$ *are*

$$\mathbb{E}\,\widehat{K}_\rho = \frac{1}{2\,\mathtt{m}_1}\,\mathrm{L}^2 + O(\mathrm{L}) \tag{20}$$

$$\mathrm{var}\,\widehat{K}_\rho = \frac{\mathtt{m}_2}{3\,\mathtt{m}_1^3}\,\mathrm{L}^3 + O(\mathrm{L}^2)\,. \tag{21}$$

Justification of the decay condition Our plan is to decompose $g_1 = h_1 + h_2$ so that h_1 will absorb the logarithmic part of g_1 at ∞ and will have a manageable Mellin transform, while Mh_2 will satisfy the decay condition by virtue of the following classical result.

Lemma 8 [16, Section 1.29] *Let* ϕ *be analytic in a sector* $-\alpha < \mathrm{Arg}\,\rho < \beta$ *with* $0 < \alpha, \beta \le \pi$*, and suppose that in this sector for some* $a < b$ *and each* $\delta > 0$

$$\phi(\rho) = O(|\rho|^{-a-\delta}) \quad \text{as} \quad |\rho| \downarrow 0\,, \ \ \phi(\rho) = O(|\rho|^{-b+\delta}) \quad \text{as} \quad |\rho| \uparrow \infty\,.$$

Then $M\phi$ *is analytic in the strip* $a < \Re\, s < b$ *and satisfies*

$$M\phi(s) = O\left(e^{-(\beta-\epsilon)\Im s}\right) \quad \text{as } \Im s \to \infty\,, \quad M\phi(s) = O\left(e^{(\alpha-\epsilon)\Im s}\right) \quad \text{as } \Im s \to -\infty\,,$$

for each $\epsilon > 0$ *uniformly in every strip strictly inside* $a < \Re\, s < b$*.*

By definition

$$g_1(\rho) = \int_0^1 f^{(1)}(\rho(1-x))(1-e^{-\rho x})\widetilde{\nu}(\mathrm{d}x)$$

which is an entire function, where

$$f^{(1)}(\rho) = e^{-\rho}\sum_{n=1}^{\infty}\frac{\rho^n}{n!}\,\mathbb{E}\,K_n$$

and the series in ρ has all coefficients positive. The same applies to the series involved in

$$1 - e^{-\rho} = e^{-\rho}\sum_{j=1}^{\infty}\frac{\rho^j}{j!}.$$

Therefore $|g_1(\rho)| \le g_1(|\rho|)$ and by (18)

$$|g_1(\rho)| = O(\mathrm{L}^3) \quad |\rho| \to \infty\,.$$

Writing (18) as

$$g_1(\rho) = \sum_{j=0}^{3} d_j \mathrm{L}^j + O(\rho^{-\epsilon})\,, \quad \rho \to \infty\,,\ \rho \in \mathbb{R}$$

and selecting

$$h_1(\rho) := d_0 e^{-1/\rho} + \log(\rho+1)\sum_{j=1}^{3} d_j \log^{j-1}\rho$$

we obtain a function $h_2 := g_1 - h_1$ which is analytic in the open halfplane $\Re\,\rho > 0$, has the same expansion in powers of L as g_1 for $\rho \to \infty$ and satisfies Lemma 8 with $a = 1$, $b = \epsilon$ and $\alpha = \beta = \pi/2$. Hence by the lemma

$$|Mh_2(s)| = O\left(e^{-(\pi/2-\delta)|s|}\right), \quad \text{for} \quad -1/2 < \Re s < \epsilon\,,\ |s| \to \infty$$

for each δ. The meromorphic continuation of Mh_1 from $-1 < \Re\,s < 0$ to the halfplane $\Re\,s > -1$ follows from the standard Mellin transform formulas

$$Me^{-1/\,\cdot}\,(s) = \Gamma(-s), \quad M\log(1+\cdot)(s) = -\Gamma(s)\Gamma(-s)\,, \quad M(\phi(\,\cdot\,)\log(\,\cdot\,))(s) = \frac{\mathrm{d}}{\mathrm{d}s}M\phi(s)$$

which by application of (14) also imply

$$|Mh_1(s)| = O\left(e^{-(\pi/2-\delta)|s|}\right), \quad \text{for} \quad -1/2 < \Re s < \epsilon\,,\ |s| \to \infty\,.$$

It follows that in a strip containing the imaginary axis $|Mg_1(s)| = O\left(e^{-(\pi/2-\delta)|s|}\right)$, and the same estimate holds for $|Mf^{(2)}(s)| = |Mg_1(s)/\Phi(-s)|$ due to Lemma 5. Thus $|Mf^{(2)}|$ decays exponentially fast and Lemma 4 is applicable.

Number of parts prior to the exponential split Let τ be an independent exponential time, with rate λ. By (9), the Mellin transform in ρ satisfies

$$M(\mathbb{E}\widehat{K}_\rho(\tau))(s) = \frac{M\widehat{\Phi}(s)}{\lambda + \Phi(-s)}.$$

The term λ in the denominator kills the zero of Φ at $s = 0$, therefore arguing as in the analysis of (13) we see that this expression has only *double* pole at $s = 0$. It follows that

$$\mathbb{E}\widehat{K}_\rho(\tau) = d\,\mathrm{L} + O(1)$$

for each $\lambda > 0$, with d depending on λ. Repeatedly appealing to (9) and Lemmas 5, 6 and 8 we obtain $f^{(m)(\rho)} = O(\mathrm{L}^m)$, which trivially implies the following rough estimate

Lemma 9 *For τ an independent exponential time, as $\rho \to \infty$ with* $\mathrm{L} = \log\rho$*,*

$$\mathbb{E}(\widehat{K}_\rho(\tau))^m = O(\mathrm{L}^m)\,, \qquad m = 1, 2, \ldots$$

Depoissonisation Modifying the argument in [10, Section 6.2] we obtain for $\rho = n \to \infty$

$$\widehat{K}_n \sim K_n\,, \quad \widehat{K}_n(t) \sim K_n(t)$$

for each $t > 0$, almost surely, with all factorial moments, and with at least two central moments. The same sandwich-type proof works because the moments grow logarithmically (thus vary regularly as $n \to \infty$).

5 Central limit theorem for K_n

We switch to fixed-n framework, in order to apply a delicate recent result due to Neininger and Rüschendorf, which we reproduce below with a minor adaptation and notational changes. Suppose a sequence (Y_n) of random variables satisfies

$$Y_n \stackrel{d}{=} b_n + Y'_{I_n}\,, \qquad n \geq 1 \tag{22}$$

where $(Y_n) \stackrel{d}{=} (Y'_n)$, the pair (b_n, I_n) is independent of (Y'_n), each I_n assumes values in $\{0, \ldots, n\}$, and $\mathbb{P}(I_n = n) < 1$ for $n \geq 1$. Let $\mu_n = \mathbb{E}\,Y_n$, $\sigma_n^2 = \mathrm{var}\,Y_n$, and $\|\cdot\|_3$ denote the $\mathcal{L}^3$-norm of a random variable.

Theorem 10 [12, Theorem 2.1] *Assume that each* $\| Y_n \|_3 < \infty$ *and that* (Y_n) *satisfies the recursion* (22). *Suppose that for some constants* $C > 0$, $\alpha > 0$ *the following three conditions all hold:*

(i)
$$\limsup_{n\to\infty} \mathbb{E} \log\left(\frac{I_n \vee 1}{n}\right) < 0\,, \qquad \sup_{n>1} \left\| \log\left(\frac{I_n \vee 1}{n}\right) \right\|_3 < \infty$$

(ii) *for some* $\lambda \in [0,\, 2\alpha[$ *and some* κ
$$\| b_n - \mu_n + \mu_{I_n} \|_3 = O(\log^\kappa n)\,, \qquad \sigma_n^2 = C\, \log^{2\alpha} n + O(\log^\lambda n)$$

(iii)
$$\alpha > \frac{1}{3} + \max\left(\kappa\,,\, \frac{\lambda}{2}\right).$$

Then the law of
$$\frac{Y_n - \mu_n}{\sqrt{C}\log^\alpha n}$$
converges weakly to the standard normal distribution.

Condition (i) says that the split I_n/n must be bounded away from 0 and 1, and the further two conditions require logarithmic growth of moments. See [12] for proof and estimates of the distance between the probability laws.

To apply the result to K_n we split the range of subordinator at an independent exponential time τ with mean 1. Let $b_n = K_n(\tau)$ be the number of parts of $\mathcal{C}_n$ produced by the multiplicative subordinator in the period between 0 to τ. Observe that (22) holds, with I_n being the number of uniform sample points larger than $\widetilde{S}_\tau$. The asymptotics of moments in Theorem 7 and Lemma 9 apply to K_n and b_n literally without change, by the virtue of depoissonisation,

Let us check now the conditions of the theorem. Since $1 \le K_n \le n$, all absolute moments of K_n are finite. Furthermore, the variable I_n/n converges strongly and with all moments to $1 - \widetilde{S}_\tau$, thus $-\log(I_n/n)$ approaches the stopped value S_τ of the *additive* subordinator, which is positive and has $\mathbb{E}S_\tau^3 < \infty$ as a consequence of $\mathtt{m}_3 < \infty$, by an easy application of the Lévy-Khintchine formula. Whence (i) is satisfied. In evaluating the asymptotics of moments we can switch to the poissonised composition with $\rho = n$, $\mathrm{L} = \log n$. Then from Theorem 7 we see that $\| \mu_n - \mu_{I_n} \|_3$ is of the order of L. By Lemma 9 also $\| b_n \|_3 = O(\mathrm{L})$, because $\mathbb{E}\, b_n^3 = O(\mathrm{L}^3)$. By the above and Theorem 7 the parameters involved in (ii) are $\alpha = 3/2$, $\lambda = 2$ and $\kappa = 1$, so condition (iii) is satisfied. Thus all conditions of Theorem 10 are fulfilled and we deduce the following conclusion:

Theorem 11 *If the Lévy measure has a continuous density on* $]0,\infty[$ *and satisfies* (L) *and* (R) *then the distribution of the random variable*
$$\frac{K_n - \mathrm{L}^2/(2\mathtt{m}_1)}{\sqrt{\mathtt{m}_2/(3\,\mathtt{m}_1^3)}\ \mathrm{L}^{3/2}}\,, \qquad \mathrm{L} = \log n$$
converges weakly to the standard normal distribution, as $n \to \infty$.

6 Small parts of the composition

We have shown in [10] that in the regular variation case all $K_{n,r}$, $r = 1, 2, \dots$ are of the same order of growth as K_n and, suitably normalised, converge almost surely to multiples of the same random variable. In the compound Poisson case these variables are bounded as n grows [7], and if the increments of (S_t) are exponentially distributed (this corresponds to the Ewens partition structure) the $K_{n,r}$'s converge to independent Poisson random variables [2]. Our next goal is proving a joint central limit theorem for the small-part counts.

Marginal central limit theorems The line of argument repeats that for K_n. Let $f_r^{(m)}(\rho)$ be the mth factorial moment of $\widehat{K}_{\rho,r}$, e.g. $f_1^{(1)}(\rho)$ is the mean number of singletons. Recall that the recursion (7) holds for $f^{(m)} = f_r^{(m)}$ with $\pi(\rho) = e^{-\rho}\rho^r/r!$. Applying the Mellin transform to the recursion we obtain

$$Mf_r^{(1)}(s) = \frac{\Gamma(r+s)}{r!}\frac{\Phi(-s:-s)}{\Phi(-s)}\,, \qquad -1 < \Re\, s < 0. \tag{23}$$

This agrees with (13) and $K_\rho = \Sigma_r\, K_{\rho,r}$ in view of the identity

$$-\Gamma(s) = \sum_{r=1}^{\infty} \frac{\Gamma(r+s)}{\Gamma(r+1)}\,, \qquad -1 < \Re\, s < 0.$$

The right-hand side of (23) can be extended meromorphically through the imaginary axis, with a *double* pole at 0, where we have the Laurent expansion

$$Mf_r^{(1)}(s) = \frac{1}{\mathtt{m}_1 r}\, s^{-2} - d_1\, s^{-1} + O(1) \tag{24}$$

where

$$d_1 = \frac{2c\mathtt{m}_1 - 2\gamma\,\mathtt{m}_1 + \mathtt{m}_2 - 2\mathtt{m}_1\,\psi^{(0)}(r)}{2\mathtt{m}_1^2\, r}$$

and $\psi^{(0)}$ is the digamma function. The decay at complex infinity is justified as before, thus by Lemma 4

$$f_r^{(1)}(\rho) = \frac{\mathrm{L}}{\mathtt{m}_1 r} + d_1 + O(\rho^{-\epsilon}). \tag{25}$$

Now we need to translate (24) into asymptotics of the right-hand side of (7) with $m = 2$. We note that evaluation of integrals with a factor of $e^{-\rho x}x^r$ essentially amounts to Tauberian-type asymptotics, since for $r > 1$ the factor x^r compensates the singularity of $\widetilde{\nu}$ at $x = 0$, and $e^{-\rho x}$ is negligible outside each fixed vicinity of 0. In particular, we have

Lemma 12 *If the condition* (L) *holds then for* $\rho \to \infty$

$$\int_0^1 \frac{e^{-\rho x}(\rho x)^r}{r!}\,\widetilde{\nu}(\mathrm{d}x) = \frac{1}{r} + O(\rho^{-\epsilon})\,.$$

Proof. Integrating by parts and replacing $\vec{\nu}$ by the right-hand side of (L), and pushing the upper integration limit to ∞ the claim is reduced to the standard integral

$$\int_0^\infty (-\log x + d)\left(\frac{e^{-\rho x}\rho^r x^{r-1}}{(r-1)!} - \frac{e^{-\rho x}\rho^{r+1}x^r}{r!}\right)\mathrm{d}x = \frac{1}{r}$$

which does not depend on d, because

$$\int_0^\infty \left(\frac{e^{-\rho x}\rho^r x^{r-1}}{(r-1)!} - \frac{e^{-\rho x}\rho^{r+1}x^r}{r!}\right)\mathrm{d}x = 0$$

as is easily checked. $\square$

Using the lemma and (25) we compute the right-hand side of (7) as

$$g(\rho) = \frac{2\mathrm{L}}{\mathtt{m}_1 r^2} + \frac{2d_1}{r} + O(1),$$

hence by Lemma 4

$$Mg(s) = \frac{2}{\mathtt{m}_1 r^2}\, s^{-2} - \frac{2d_1}{r}\, s^{-1} + O(1)\,.$$

We compute then

$$Mf_r^{(2)}(s) = \frac{Mg(s)}{\Phi(-s)} = -\frac{2}{\mathtt{m}_1^2 r^2}\, s^{-3} + \left(\frac{\mathtt{m}_2}{r^2 {\mathtt{m}_1}^3} + \frac{2d_1}{r\mathtt{m}_1}\right) s^{-2} + O(s^{-1})$$

which yields again by Lemma 4

$$f_r^{(2)}(\rho) = \frac{1}{r^2 \mathtt{m}_1^2}\mathrm{L}^2 + \left(\frac{\mathtt{m}_2}{r^2 {\mathtt{m}_1}^3} + \frac{2d_1}{r\mathtt{m}_1}\right)\mathrm{L} + O(1)$$

(the decay condition is again checked by application of Lemma 5). This together with (25) implies

$$\mathrm{var}\, K_{r,\rho} = \left(\frac{\mathtt{m}_2}{r^2 \mathtt{m}_1^3} + \frac{1}{r\mathtt{m}_1}\right)\mathrm{L} + O(1)\,.$$

Decomposing as in (8) at rate 1 exponential time, the Mellin transform of the expectation of $b_\rho = \widehat{K}_{\rho,r}(\tau)$ has $\Phi(-s)+1$ in the denominator, hence all moments of b_ρ remain bounded as $\rho \to \infty$. Passing to the fixed-n version and applying Theorem 10 with $\alpha = 1/2, \lambda = \kappa = 0$ we obtain

Theorem 13 *Under our assumptions on the Lévy measure, for large n the distribution of each $K_{n,r}$ is approximately normal, with moments*

$$\mathbb{E}K_{n,r} = \frac{1}{\mathtt{m}_1 r}\log n\, + O(1), \qquad \mathrm{var}\, K_{n,r} = \left(\frac{\mathtt{m}_2}{r^2 \mathtt{m}_1^3} + \frac{1}{r\mathtt{m}_1}\right)\log n + O(1)\,.$$

The same is true for $\widehat{K}_{\rho,r}$ as $\rho \to \infty$.

Joint central limit theorem Our strategy to prove a joint central limit theorem for the small parts counts is to consider more general finite patterns and functionals of composition such as $\Sigma_r\, a_r \widehat{K}_{\rho,r}$. By linearity, each variable of this kind decomposes as in (8), hence the method we applied to $\widehat{K}_{\rho,r}$ can be used to justify the normal limit. As an instance of such a functional consider

$$\widehat{K}_{\rho,[r]} := \widehat{K}_{\rho,1} + \cdots + \widehat{K}_{\rho,r}$$

and define $f_{[r]}^{(m)}(\rho)$ to be the mth factorial moment of $\widehat{K}_{\rho,[r]}$. The corresponding pattern is $\{1,\ldots,r\}$, thus the moments satisfy (7) with $\pi(\rho) = \sum_{j=1}^r e^{-\rho}\rho^j/j!$. Obviously from (25)

$$f_{[r]}^{(1)} = \frac{h_r}{\mathtt{m}_1}\mathrm{L} + O(1)$$

where $h_r := \sum_{j=1}^r 1/j$ are the harmonic numbers. Letting Mellin's machine roll to produce $f_{[r]}^{(1)} \to g \to Mg \to Mf_{[r]}^{(2)} \to f_{[r]}^{(2)}$ the variance is computed as

$$\mathrm{var}\, K_{\rho,[r]} = \left(\frac{\mathtt{m}_2 h_r^2}{\mathtt{m}_1^3} + \frac{h_r}{\mathtt{m}_1}\right)\mathrm{L} + O(1).$$

Similar computation for the pattern $E = \{i, j\}$ and Theorem 13 yield the covariance

$$\mathrm{cov}(\widehat{K}_{\rho,i}, \widehat{K}_{\rho,j}) = \left(\frac{\mathtt{m}_2}{\mathtt{m}_1^3} \frac{1}{i\,j} + 1(i=j) \frac{1}{j\,\mathtt{m}_1} \right) \mathrm{L} + O(1)\,.$$

Theorem 14 *Under our assumptions on ν, as $n \to \infty$, the infinite random sequence*

$$\big((K_{n,r} - \mathbb{E}K_{n,r}) \log^{-1/2} n\,, \quad r = 1, 2, \ldots, n \big)$$

converges in law to a multivariate gaussian sequence with the covariance matrix

$$\left(\frac{\mathtt{m}_2}{\mathtt{m}_1^3} \frac{1}{i\,j} + 1(i=j) \frac{1}{j\,\mathtt{m}_1} \right)_{i,j=1}^{\infty}\,. \tag{26}$$

Proof. To justify the joint gaussian law it suffices to establish convergence to a gaussian limit for each finite linear combination

$$Q_n = \sum_r a_r K_{n,r}\,.$$

From the above facts about $K_{n,r}$ and from (26) it follows that both the expectation and the variance of Q_n grow like L. Also, due to an obvious additivity, Q_n satisfies a distributional equation of the type (22) provided we split the range of subordinator by the value at an independent exponential time τ. Then b_n is equal to the contribution to Q_n by the jumps of subordinator before τ. The conditions of Theorem 10 are checked exactly as for K_n or $K_{n,r}$, thus by this theorem Q_n is approximately gaussian. □

7 The gamma case and further examples

In the case of the gamma subordinators we have

$$c = -\log\theta\,, \qquad \mathtt{m}_j = \frac{(j-1)!}{\theta^j}\,,$$

and the singular expansion

$$\Phi(-s:-s) = -s^{-1} + (c-\gamma) - d_2 s + O(s^2)$$

with constant

$$d_2 = \int_0^1 \log x \,\frac{x(1-x)^{\theta-1} + \log(1-x)}{-x\log(1-x)}\,\mathrm{d}x$$

which does not seem to simplify. The Mellin transform of $\mathbb{E}\widehat{K}_\rho$ is given by the formula

$$Mf^{(1)}(s) = \frac{-\Gamma(s)}{\log(1-s/\theta)} \int_0^1 \frac{x^{-s}(1-x)^{\theta-1}}{-\log(1-x)}\,\mathrm{d}x$$

which defines a function which is meromorphic for $\Re\, s > -1$ with a sole singularity at $s = 0$. The Laurent expansion of $Mf^{(1)}$ at $s = 0$ involves

$$\mathtt{m}_j = \frac{(j-1)!}{\theta^j}\,, \quad c = -\log\theta\,, \quad d_1 = \frac{\gamma^2}{2} + \frac{\pi^2}{12}\,.$$

Leaving only principal terms, the asymptotics of moments is

$$\mathbb{E}\widehat{K}_\rho = \frac{\theta}{2}\,\mathrm{L}^2 + O(\mathrm{L})\,, \qquad \mathrm{var}\,\widehat{K}_\rho = \frac{\theta}{3}\,\mathrm{L}^3 + O(\mathrm{L}^2)$$

as was promised in Theorem 1.

Further examples Another instance of a gamma-type subordinator was introduced in [8] to describe a composition resembling the ordered Ewens sampling formula. This subordinator has (multiplicative) Lévy measure

$$\widetilde{\nu}(\mathrm{d}x) = (1-x)^{\theta-1}x^{-1}\,\mathrm{d}x\,, \qquad x \in]0,1]$$

with parameter $\theta > 0$. The Laplace exponent given by the formula

$$\Phi(s) = \sum_{j=1}^{\infty}\left(\frac{1}{j+\theta-1} - \frac{1}{j+\theta-1+s}\right)$$

is a function interpolating the generalised harmonic numbers

$$\Phi(n) = \sum_{j=1}^{n}\frac{1}{j+\theta-1}\,, \tag{27}$$

which can be seen as a combinatorial analogue of the logarithmical Laplace exponent for the gamma subordinator.

The basic characteristics of subordinator are readily expressed in terms of the polygamma function

$$\psi^{(k)}(\theta) = \frac{\mathrm{d}^{k+1}\,\log\Gamma(\theta)}{\mathrm{d}\,\theta^{k+1}}.$$

Thus we have

$$\vec{\nu}(x) = -\log x - \psi^{(0)}(\theta) + O(x^{-1})$$

as is shown by expanding

$$\vec{\nu}(x) = \int_x^1 (1-z)^{\theta-1}z^{-1}\,\mathrm{d}z = -\log x + c(\theta) - \gamma + c_1(\theta)x + \dots\,,$$

substituting $\theta = 1$ to see that $c(1) = \gamma = -\psi^{(0)}(1)$, then differentiating in θ and sending $x \to 0$ to obtain

$$c'(\theta) = \int_0^1 (1-z)^{\theta-1}z^{-1}\log(1-z)\,\mathrm{d}z = -\psi^{(1)}(\theta)\,, \tag{28}$$

whence $c(\theta) = -\psi^{(0)}(\theta)$. It follows as in Appendix or directly from (27) that the expansion at infinity is

$$\Phi(\rho) = \mathrm{L} - \psi^{(0)}(\theta) + O(\rho^{-1})\,.$$

The moments are computed by further differentiating (28) as

$$\mathtt{m}_j = \int_0^1 (1-x)^{\theta-1}x^{-1}|\log(1-x)|^j\,\mathrm{d}x = (-1)^{j+1}\psi^{(j)}(\theta).$$

See [3] for computation of some densities related to this family of subordinators, and [15] for further examples of Lévy measures with logarithmic singularity.

Oscillatory asymptotics for a discrete measure The constant term c assumed in (L) cancels in the asymptotics of the moments. However, this assumption is essential by our approach. Let us assume a bounded oscillating term in place of c, and examine how the asymptotics could be affected. For ease of computation we consider the atomic measure

$$\widetilde{\nu}(\mathrm{d}x) = \sum_{j=1}^{\infty}\delta_{1/e^j}(\mathrm{d}x)\,.$$

For this measure $\Phi(n)$ is equal to the expected maximum in a sample of n geometric random variables, which was analysed in [5, Example 12].

Denoting $\lfloor \cdot \rfloor$, respectively $\{ \cdot \}$, the integer and the fractional parts of a positive number, we have

$$\vec{\nu}(x) = \lfloor -\log x \rfloor = -\log x - \{-\log x\}\,, \qquad x \in]0,1]$$

thus there is a logarithmic singularity but the expansion (L) does not hold, because the second term oscillates between 0 and 1. Condition (R) is satisfied since $\vec{\nu}(x) = 0$ for $x > 1/e$ and all moments are finite,

$$\mathtt{m}_k = \sum_{j=1}^{\infty} |\log(1 - e^{-j})|^k\,,$$

e.g. $\mathtt{m}_1 = 0.6843\,,\ \mathtt{m}_2 = 0.2345$ (truncated at four decimals). Furthermore,

$$\widehat{\Phi}(\rho) = \sum_{j=1}^{\infty}(1 - e^{-\rho/e^k})\,, \quad M\widehat{\Phi}(s) = \frac{-\Gamma(s)}{1-e^s}\,, \quad \Phi(s) = \sum_{j=1}^{\infty}(1-(1-e^{-j})^s).$$

It is seen that $M\widehat{\Phi}$ has a *double* pole at $s = 0$ with Laurent expansion

$$M\widehat{\Phi}(s) = \frac{1}{s^2} - \frac{\gamma + 1/2}{s} + \frac{1}{2}(1 + 6\gamma + 6\gamma^2 + \pi^2) + O(s)$$

and infinitely many *simple* poles on the imaginary axis at

$$s_k = 2\pi\mathtt{i}k\,, \quad k \in \mathbb{Z} \setminus \{0\}.$$

The conclusion of Lemma 4 still holds, which can be justified by applying the inverse Mellin transform formula and integrating over increasing rectangular contours, as $k \to \infty$, with sides $\Im s = 2\pi k + 1/2$, $\Im s = -2\pi k - 1/2$, $\Re s = 1/2$, $\Re s = -1/2$. Thus

$$\widehat{\Phi}(\rho) = \mathrm{L} + \gamma + 1/2 + \phi(\mathrm{L}) + O(\rho^{-1+\epsilon})\,, \qquad \rho \to \infty \tag{29}$$

where the contribution of the imaginary poles amounts to the term $\phi(\mathrm{L})$ given by the formula

$$\phi(u) = -\sum_{k \in \mathbb{Z}\setminus\{0\}} \Gamma(2\pi\mathtt{i}k)\, e^{2\pi\mathtt{i}ku}\,,$$

which is a periodic function with period one and a tiny amplitude. The fluctuations of the $O(1)$ term in (29) are not asymptotically negligible, though they are very small due to the fast decay of the Fourier coefficients, e.g.

$$\Gamma(s_1) = \overline{\Gamma(s_{-1})} = (0.126 + 0.501\,\mathtt{i})\,10^{-4}.$$

(truncated at seven decimals).

It follows that $Mf^{(1)}$ has a triple pole at 0 and simple imaginary poles $s_k = 2\pi\mathtt{i}k$, hence

$$f^{(1)} = \frac{1}{2\mathtt{m}_1}\mathrm{L}^2 + \left(\frac{\gamma + 1/2}{\mathtt{m}_1} + \frac{\mathtt{m}_2}{2\mathtt{m}_1^2}\right)\mathrm{L} + \phi_1(\mathrm{L}) + O(\rho^{-1+\epsilon})$$

where ϕ_1 is another periodic function. Thus oscillation prevails only in the third term in the expansion of $f^{(1)}$. However, computing g_1, as in Lemma 6, we should include $\mathrm{L}^2\widehat{\Phi}(\rho)$, thus arriving at

$$g_1(\rho) = \frac{1}{\mathtt{m}_1}\mathrm{L}^3 + \phi_2(\mathrm{L})\mathrm{L}^2 + \ldots$$

with oscillating ϕ_2. As a consequence we will have

$$f^{(2)} = \frac{1}{4\mathtt{m}_1^2}\mathrm{L}^4 + \phi_3(\mathrm{L})\mathrm{L}^3 + \ldots,$$

and this suggests that the principal $O(\mathrm{L}^3)$-term in the asymptotic expansion of the variance $\operatorname{var}\widehat{K}_\rho$ will oscillate, though we could not establish this fact rigorously.

In a similar situation of sampling from the geometric distribution, the expectation of the number of different values in a sample also involves an oscillating second term, and the same is true for the factorial moment of order 2 [14]. However, this has no impact on the principal asymptotics of the variance: the oscillating terms cancel and the variance converges to a constant, as had been shown long ago by Karlin, see [11, Example 6, p. 385].

8 Appendix

Lemma 15 *The four conditions in* (L) *are equivalent.*

Proof. The equivalence of expansions of ν and $\widetilde{\nu}$ follows by the change of variables $x = 1 - e^{-y}$ and $y = -\log(1-x)$. For example, assuming the expansion of ν we obtain that of $\widetilde{\nu}$ by using

$$\widetilde{\nu}[x,1] = \nu[-\log(1-x), \infty]$$

and substituting

$$-\log(-\log(1-x)) = -\log(x(1+x/2+\ldots)) = -\log x + O(x).$$

The expansions of integrals are equivalent to the expansions of measures by writing

$$\Phi(\rho)/\rho = \int_0^\infty e^{-\rho y}\nu[y,\infty]\,\mathrm{d}y\,, \qquad \widehat{\Phi}(\rho)/\rho = \int_0^1 e^{-\rho x}\widetilde{\nu}[x,1]\,\mathrm{d}x$$

and using the classical integral

$$\rho\int_0^\infty e^{-\rho y}\log(1/y)\,\mathrm{d}y = \mathrm{L} + \gamma$$

together with standard properties of the Laplace transform. □

Lemma 16 *The function*

$$\Phi(s) = \int_0^1 (1-(1-x)^s)\,\widetilde{\nu}(\mathrm{d}x)$$

has no zeros for $\mathrm{Re}\, s > 0$*. And if* $\Phi(s) = 0$ *for purely imaginary* s*, then* $\widetilde{\nu}$ *is atomic, with support* $\{1-a^k, k=1,2,\ldots\}$ *for some* $0 < a < 1$*.*

Proof. For $\Re\, s > 0$ and $x \in]0,1[$ we have $|(1-x)^s| < 1$. Therefore $\Re\,(1-(1-x)^s) > 0$ and the integral cannot be zero. For real $r \neq 0$ the equality $\Phi(\mathrm{i}\, r) = 0$ is only possible when $1 = (1-x)^{\mathrm{i}\, r}$ holds $\widetilde{\nu}$-almost everywhere, but such x is of the form $x = 1 - \exp(-2\pi k/|r|)$ for some $k = 1, 2, \ldots$. □

References

[1] J.A. Armstrong and N. Bleistein, Asymptotic expansions of integrals with oscillatory kernels and logarithmic singularities, *SIAM J. Math. Anal.*, 11: 300–307, 1980.

[2] R. Arratia, A.G. Barbour and S. Tavaré *Logarithmic combinatorial structures: A probabilistic approach* 2002 (forthcoming book)

[3] C. Berg and A.J. Duran Some transformations of Hausdorff moment sequences and harmonic numbers, preprint, 2004.

[4] M.V. Fedoryuk, *Asymptotics: Integrals and series.* Nauka, Moscow, 1987.

[5] P. Flajolet, X. Gourdon and P. Dumas, Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.* 144: 3–58, 1995.

[6] A.V. Gnedin, The representation of composition structures, *Ann. Probab.* 25: 1437-1450, 1997.

[7] A.V. Gnedin. The Bernoulli sieve, *Bernoulli* 10: 79-96, 2004.

[8] A.V. Gnedin. Three sampling formulas, *Combinatorics, Probability and Computing* 13: 185-193, 2004.

[9] A.V. Gnedin and J. Pitman. Regenerative composition structures. *Ann. Prob.* (to appear)

[10] A.V. Gnedin, J. Pitman and M. Yor. Asymptotic laws for compositions derived from transformed subordinators. (available at arXiv:math.PR/0403438)

[11] S. Karlin. Central limit theorems for certain infinite urn schemes. *J. Math. Mech.*, 17:373–401, 1967.

[12] R. Neininger and L. Rüschendorf, On the contraction method with degenerate limit equation. *Ann. Prob.* (to appear)

[13] J. Pitman. Combinatorial stochastic processes. Technical Report 621, Dept. Statistics, U.C. Berkeley, 2002. Lecture notes for St. Flour course, July 2002. (available via www.stat.berkeley.edu)

[14] H. Prodinger. Compositions and patricia tries: no fluctuations in the variance! preprint 2003.

[15] J. Pitman and M. Yor. Infinitely divisible laws associated with hyperbolic functions. *Canad. J. Math.*, 55(581):292–330, 2003.

[16] E.C. Titchmarsh *Introduction to the theory of Fourier integrals,* Oxford Univ. Press 1937.